

Common Data Analysis Preservation Platform for LHC Experiments?

Tibor Šimko & Sünje Dallmeier-Tiessen
CERN

AAHEP7
Stony Brook University, New York
2 April 2014

1 “Small Data”

2 “Big Data”

3 Common Platform

4 Conclusions

1

“Small Data”

Examples of Current Practices

CDS

Restricted Data

CERN Accelerating science Sign in Directory

CERN Document Server

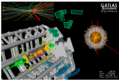
Search Submit Help Personalize

Home » Event displays of a Higgs to 4 μ candidate event » Access to Fulltext

Information Discussion Files

Event displays of a Higgs to 4 μ candidate event - ATLAS Collaboration - ATLAS-EVENT-DISPLAY-2012-005


Main file(s):

 fig_33

[fig_33.png](#) [1.94 MB] 31 Oct 2011, 14:23
[fig_33.png \(icon-1440\)](#) [940.97 KB] 31 Oct 2011, 14:23
version 1 [fig_33.png \(icon-180\)](#) [33.6 KB] 31 Oct 2011, 14:23
[fig_33.png \(icon-640\)](#) [240.33 KB] 31 Oct 2011, 14:23


Additional file(s):

Additional file(s):

 fig_32

version 1 [fig_32.png](#) [3.57 MB] 31 Oct 2011, 14:13 *Persint display*


RESTRICTED

 run204769_evt71902630

version 1 [run204769_evt71902630.png](#) [5.56 MB] 31 Oct 2011, 14:13 *Poster*

Press file(s):

RESTRICTED

 fig_33 notlogo

DEV

CERN Document Server

Search

Submit

Help

Personalize

Home > Articles & Preprints > Published Articles > First observation of $\Lambda_b^0 \rightarrow J/\psi K^+ K^-$ and search for $\Lambda_b^0 \rightarrow J/\psi \phi$ decays > Access to Fulltext

Information

Discussion

Files

Plots

Linkbacks

First observation of $\Lambda_b^0 \rightarrow J/\psi K^+ K^-$ and search for $\Lambda_b^0 \rightarrow J/\psi \phi$ decays - Aaij, R *et al* -
arXiv:1308.5916

Main file(s):

 prd.88.e072005

version 1 prd.88.e072005.pdf [1.32 MB] 07 Nov 2013, 15:17 APS Open Access article

Additional file(s):

 Related data file(s)

version 1 Related data file(s).zip [10.36 MB] 02 Sep 2013, 16:41

arXiv file(s):

 arXiv:1308.5916version 4
arXiv:1308.5916.pdf [4.36 MB] 26 Oct 2013, 04:15
(see previous)

Similar records

INSPIRE

Public Data

Information Citations (0) Files

Additional data from: First observation of $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ and search for $\bar{B}^0 \rightarrow J/\psi \phi$ decays

LHCb Collaboration

Description: The first observation of the $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ decay is presented with a data sample corresponding to an integrated luminosity of 1.0 fb^{-1} of pp collisions at a center-of-mass energy of 7 TeV collected with the LHCb detector. The branching fraction is measured to be $\mathcal{B}(\bar{B}^0 \rightarrow J/\psi K^+ K^-) = (2.53 \pm 0.31 \pm 0.19) \times 10^{-6}$, where the first uncertainty is statistical and the second is systematic. An amplitude analysis of the final state in the $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ decay is performed to separate resonant and nonresonant contributions in the $K^+ K^-$ spectrum. Evidence of the $a_0(980)$ resonance is reported with statistical significance of 3.9 standard deviations. The corresponding product branching fraction is measured to be $\mathcal{B}(\bar{B}^0 \rightarrow J/\psi a_0(980), a_0(980) \rightarrow K^+ K^-) = (4.70 \pm 3.31 \pm 0.72) \times 10^{-7}$, yielding an upper limit of $\mathcal{B}(\bar{B}^0 \rightarrow J/\psi a_0(980), a_0(980) \rightarrow K^+ K^-) < 9.0 \times 10^{-7}$ at 90 % confidence level. No evidence of the resonant decay $\bar{B}^0 \rightarrow J/\psi \phi$ is found, and an upper limit on its branching fraction is set to be $\mathcal{B}(\bar{B}^0 \rightarrow J/\psi \phi) < 1.9 \times 10^{-7}$ at 90 % confidence level. 001595651 542__

Note: * Temporary entry *

This dataset complements the following publication:

[First observation of \$\bar{B}^0 \rightarrow J/\psi K^+ K^-\$ and search for \$\bar{B}^0 \rightarrow J/\psi \phi\$ decays](#)

Record created 2014-01-27, last modified 2014-01-27

Information Citations (3) Files

Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHCATLAS Collaboration (Aad, Georges (Freiburg U.) [...]) [Show all 2923 authors](#)Cite as: ATLAS Collaboration (2013) HepData, <http://doi.org/10.7484/INSPIREHEP.DATA.RF5P.6M3K>

Description: -2 log Likelihood for the $H \rightarrow ZZ^* \rightarrow 4l$ channel in the $(\mu_{ggF+ttH} \cdot B/BSM, \mu_{VBF+VH} \cdot B/BSM)$ plane for a Higgs boson mass $m_H = 125.5$ GeV.

Preview not available

Note: * Temporary entry *

This dataset complements the following publication:

[Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC](#)

Record created 2013-09-11, last modified 2013-12-16

⇒ Export
[BibTeX](#), [EndNote](#), [LaTeX\(US\)](#), [LaTeX\(EU\)](#), [Harvmap](#), [MARC](#),
[MARCXML](#), [NLM](#), [DC](#)

Information References Citations Files Plots HepData

Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment - ATLAS Collaboration (Aad, Georges et al.) JHEP 1310 (2013) 130 arXiv:1308.1841 [hep-ex] CERN-PH-EP-2013-110

THIS DATA COMES FROM DURHAM HEPDATA PROJECT

DATASETS:

Description: MET/sqrt(HT) distributions for the multi-jet + flavour stream with PTmin=50 GeV, and exactly eight jets, with the signal region selection, other than that on MET/sqrt(HT) itself. A selection of zero b-jets is applied.

[Go to the record](#)

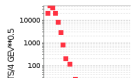
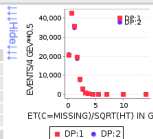
Plain

$|ETARAP(B^-jet)| < 2.5$
 $N - BJETS(p_T > 40 GeV) = 0.0$
 $N - JETS(p_T > 50 GeV) = 8.0$
 $pp \rightarrow JETS MM$
 $= dATA = MC$
 $|ETARAP(jet)| < 2.0$

$ET_{MISSING}/\sqrt{HT} (GeV^{0.5})$

EVENTS/4 GeV^{0.5}

$ET_{MISSING}/\sqrt{HT} (GeV^{0.5})$	EVENTS/4 GeV ^{0.5}	EVENTS/4 GeV ^{0.5}
0.25	20504.0	21017.74 ± 0.25
0.75	42632.0	42806.83 ± 0.00
1.25	35848.0	35159.43 ± 0.00
1.75	19376.0	18926.27 ± 0.00
2.25	7872.0	7742.63 ± 0.00
2.75	2720.0	2686.48 ± 1.43
3.25	792.0	880.64 ± 4.83
3.75	100.0	302.77 ± 0.73



Invenio

Common Technology

- mature **digital repository** platform, originated in 2002
 - articles, books, notes, photos, videos, software, data
- Invenio-based services at CERN:

CERN Document Server
Access articles, reports and multimedia content in HEP



- co-developed by an international collaboration:

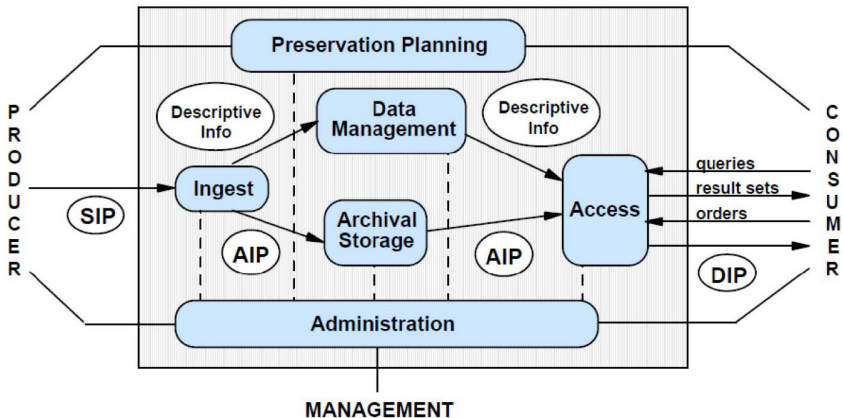


- participating in and collaborating with EC and non-EC projects:



BlogForever

OAIS Recommended Preservation Practices



SIP = Submission Information Package · AIP = Archival Information Package · DIP = Dissemination Information Package

```

/tmp> unzip -t record_59_v1.zip
Archive: record_59_v1.zip
testing: bag-info.txt           OK
testing: manifest-md5.txt      OK
testing: bagit.txt             OK
testing: data/BlogImage_2.metadata OK
testing: data/BlogPostPage_page.html OK
testing: data/BlogPostPage_page.metadata OK
testing: data/BlogImageThumbnail_3.jpeg OK
testing: data/BlogImageThumbnail_3.metadata OK
testing: data/Post_2013-08-20_14:21:21ucnxAN.metadata OK
testing: data/Post_2013-08-20_14:21:21ucnxAN.xml1_mets OK
testing: data/record_59_v1.xml OK
testing: data/BlogImage_2.jpeg OK
No errors detected in compressed data of record_59_v1.zip.
/tmp>

```

Post from blog: Tech Talk

 gadgetcscn

 Posted date

2012/02/28

 Archived date

2013/08/20

 Topic

No topic yet

 Tags

[Intel](#) [Mobiles](#) [samsung](#)

[world mobile congress](#)

Add your own tags

 License

Unknown

div.bfe-blog-navigation 270px x 50px

 Export as

BibTeX, MARC, MARCXML, METS, DC, PDF, JPEG

 Share



 Other options

Add to Personal Collection
 Edit This Record
 Ask for Deletion
 Manage Files of This Record
 Download BagIt
 Similar Posts

Intel's mobile plan: Zero to Hero



Intel CEO Paul Otellini show inoff Orange new phone at MWC

The hottest smartphone maker at this year's Mobile World Congress in Barcelona isn't

You didn't know Intel designed phones? Neither did most people until the chipmaking big designed a top-of-the-line smartphone for the global telecom giant Orange.

It all started about a year ago, when extremely frustrated top brass at Intel decided the c its chips would work on mobile phones — a field dominated by mobile chip designing rtr PowerPoint presentations, no matter how impressive, just weren't persuading phone ma

So Intel hired several hundred engineers from all over Silicon Valley and opened up a sta inside the company.

Late last year, Intel began to show off the fruit of its labors to the same folks who used b device — designed entirely and exclusively at Intel's Santa Clara headquarters — has al runs Google's (GOOG, Fortune 500) Android operating system beautifully, and looks rer

That shouldn't be surprising. Intel's head of mobility, Mike Bell, worked at Apple (AAPL, iPhone project before leaving for Palm.

"Our device is fully buzzword-compliant," said Bell, in an interview on Tuesday. "Intel is becoming handheld computers. It's natural for us to get into this space."

Intel finally ups the ante in mobile

Intel built a few thousand of the phones and started handing them out to any and all pote PowerPoint couldn't. In addition to Orange, Indian carrier Lava is planning on shipping a quarter, and Lenovo launched a similar phone last month.

ZENODO

Long-Tail Sciences

zenodo

Research. Shared.

[✉](#) [Search](#) [Communities](#) [Browse ▾](#) [Upload](#) [Get started ▾](#) [Administration ▾](#)[tibor.simko@cern.ch ▾](#)

Search 1193 records for

[Q Search](#)

Any Collection

[Publications](#) (883)[Presentations](#) (179)[Datasets](#) (46)[Posters](#) (45)[Software](#) (23)[More ...](#)

Any Author

[Aiftimiei, Cristina](#) (18)[De Angelis, Filip...](#) (18)[Alves, Everton ...](#) (16)[Fuhrmann, Patr...](#) (16)[Gazetas, George](#) (16)[More ...](#)

Showing records 1 to 10 out of 1193 results.

1

[25 January 2014](#) [Presentation](#) [Open access](#)**The Effective Use of "Research Tools" and Resources – Training of Trainers (TOT)**[View](#)[Ale Ebrahim, Nader](#)

With the increasing use of information and communications technology (ICT), researchers are able to use computer software tools to find, organize, analyze, and share relevant information. However, there are hundreds of such tools to select from, for ...

Uploaded by [Nader Ale Ebrahim](#) on 28 January 2014.

INVENIO » “Code You Can Cite”

- built ZENODO ↔ GitHub bridge
- archive software automatically upon release; mint it with a DOI

zenodo Research, Shared.

Get started

- 1 Flip the switch
- 2 Create a release
- 3 Get the badge

invenio-cery/decouple

ON

invenio-cery / decouple

v1.1.3

Releases

```
{
  "name": "invenio, invenio",
  "affiliation": "Institut f\u00fcr Theoretische Ph",
  "description": "This repository contains the soft",
  "access_right": "open",
  "license": "MIT-license",
  "related_identifiers": [
    {
      "identifier": "arXiv:1401.0088",
      "relation": "isCitedBy"
    }
  ]
}
```

.zenodo.json

DOI: 10.5281/zenodo.8345

INVENIO » Code ↔ Data ↔ Paper

- link code to data, link code to papers
- first use cases: hep-ex/0011057, arXiv:1401.0080

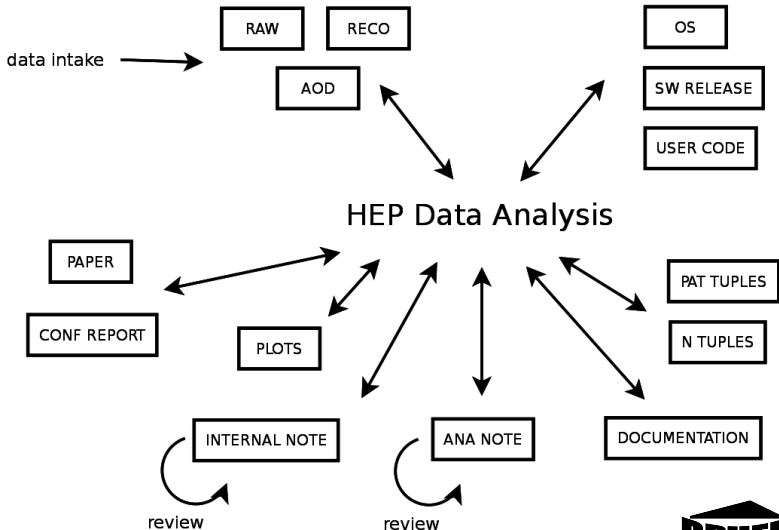
The image displays two overlapping web interfaces. The background interface is Zenodo, showing a record for 'decouple software associated to arXiv:1401.0080'. The foreground interface is INSPIRE, showing a record for the paper 'A Novel Approach to Higgs Coupling Measurements' by Kyle Cranmer, Sven Heinke, and David Lopez-Fey. The INSPIRE record includes a description of the software, a link to the Zenodo repository, and a list of keywords including 'Automatic Exports', 'inspire', 'Higgs', 'CMS/LHC/Gol', 'new physics', and 'decouple'. The paper record also features three plots showing confidence contours in the μ - κ plane.

2

“Big Data”

Data Analysis Preservation Pilot

» Preserve an Analysis?



■ CMS

- WW cross section & Higgs search
- EWK-10-009 (PAS)
- HIG-10-003 (PAS)

■ LHCb

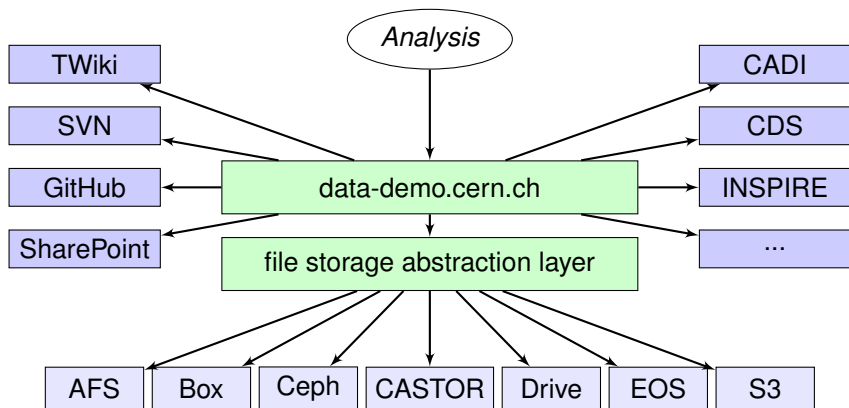
- $B_s^0 \rightarrow J/\psi K_S^0 0$
- arXiv:1205.0934

■ ALICE

- just started

3

Common Platform



Data Analysis Preservation Demo



Data Analysis Preservation Platform Demo :: Search :: Submit :: Personalize :: Help

Powered by invenio v1.1.2.1211-9359eb-dirty

Maintained by tibor.simko@cern.ch

This site is also available in the following languages:

Afrikaans العربية български Català Český Deutsch Ελληνικά English Español
فارسی Français Hrvatski Galego ქართული Italiano Kinyarwanda Lietuvių Magyar
日本語 Norsk/Bokmål Polski Português Română Русский Slovensky Svenska
Українська 中文(D) 中文(繁)

CMS Data Analysis Mock-Up [DO NOT SUBMIT YET]

NOTE: Access to all submitted data will be restricted to the CMS collaboration only.

Basic Info



Physics Info



Pre-selection step



Custom analysis step?



Final selection step



Internal Documentation



Internal Discussion



Final selection step

OS

Analysis software

User code

Example of supported repositories:
git@github.com: johndoe/myrepo.git
svn@svnweb.cern.ch:cern/wsvn/myrepo

Harvest? yes, harvest user code
 no, create link only

Input Data Files taken from output of pre-selection step
 taken from output of custom analysis step

Output Data Files [+ Add another file](#)

Example of supported protocols:
xroot://castorpublic.cern.ch/castor/cern.ch/user/jfjohndoe/mydir/myfile.root
root://eospublic.cern.ch/eos/lhcb/.../myfile.root
file:///tmp/myfile.root
http://john.doe.example.org/myfile.root

Harvest? yes, harvest files no, create link only

4

Conclusions

- reaching out from “small data” towards “big data”
 - added software
 - adding data file abstraction layer
 - adding visualisation
- data analysis preservation platform
 - pilot exercise under way
 - CMS, LHCb, ALICE actively engaged
 - offering internal services to collaborations
 - offering automated feeds to CDS, INSPIRE, etc
- CERN became DataCite member

*“Capture comprehensive metadata information
to enable future data reuse”*